

Bírálat Do Van Tien

*Multi-Server Markov Queueing Models: Computational
Algorithms and ICT Applications*
című MTA doktori disszertációjáról.

Témaválasztás

Do Van Tien MTA doktori fokozat megszerzéséhez a "Többserveres Markovi kiszolgálási modellek: számítási algoritmusok és infokommunikációs alkalmazások" című disszertációval pályázott. A jelölt az elmúlt kb. egy/másfél évtizedben ICT rendszerek teljesítmény elemzése és kiértékelése területén fejtett ki jelentős tudományos tevékenységet, amelyben elért eredményeit fent említett dolgozatában ismerteti. Ez a tématerület rendkívül intenzíven veti fel a kutatási problémákat, szorosan kapcsolódva a technológiai fejlesztésekhez. Pont az elmúlt másfél évtizedben jelentek meg és (részben) terjedtek el olyan vezetékes és vezeték nélküli hálózati átviteli technológiák, amelyekben a kiszogálandó "ügyfelek" száma és a nyújtandó szolgáltatások komplexitása drámai mértékben megnőtt. A tömegkiszolgálási modellek tehát alapvető fontosságúak ilyen rendszerek elemzésében és kiértékelésében, sőt az alkalmazandó sorbanállási modellek és a megoldásukhoz szükséges számítási eljárások állandó továbbfejlesztése/megújítása szükséges. Példaként (természetesen a teljesség igénye nélkül) gondoljunk arra, hogy az elmúlt 15 évben mennyit fejlődött a mobil kommunikáció, milyen mértékben növekedett a web szerverek száma és teljesítménye, megjelent és elterjedt a többprotokollós cimkekapcsolás (MPLS), az optikai kapcsolat óriási mértékben haladt előre, és a virtualizáció a jelen és jövő ICT rendszereiben alapvető koncepcionális megközelítés lesz. Megállapítható tehát, hogy a jelölt a gyakorlat szempontjából is releváns kérdések megválaszolásához végzett elméleti kutatásokat.

A fentiekkel összhangban, a jelölt három fő témában ismerteti eredményeit a disszertációban, nevezetesen:

- Többserveres Markovi sorbanállási rendszer általánosítása és alkalmazásai
- Az újrapróbálkozási sorokkal modellezhető rendszerek jellemzőinek hatékony meghatározása
- A CPP/M/c vakációs (szerver megszakításos) sorbanállási modell

A disszertáció és a tézisek értékelése

Még mielőtt rátérnék a részletes értékelésre, szeretném előrebecsajítani, hogy a disszertációt és téziszfüzetet formailag és tartalmilag is mindenképpen alkalmasnak tartom arra, hogy ezek alapján az MTA doktori cím odaítélését meg lehessen ítélni.

A dolgozat 5 részből és 8 fejezetből áll. A disszertáció felépítése és az eredmények prezentálása az egyes fejezetekben egy jól kigondolt struktúra következetes megvalósítása mentén történik. Az eredményeket ismertető (lényegi) részek közül a negyedik a második és harmadik részekhez képest sokkal kisebb kiméretű, de ez azt hiszem annak fényében nem probléma, hogy a részek az egyes nagyobb témaköröket hivatottak elsősorban elválasztani. Mindazonáltal szembeötlő, hogy a negyedik rész egyetlen fejezete sokkal rövidebb, mint a többi részben található fejezetek többsége.

A jelölt három tézisben foglalja össze tudományos eredményeit és ezek a tézisek egyértelműen megfeleltethetők a "lényegi" részeknek. Az altézisekben mindenhol hivatkozik a szerző az adott fejezetre, illetve a fejezetekben szereplő tételekre, formulákra, ábrákra. A téziszfüzet tehát ilyen szempontból nem tekinthető "önjárónak" (a téziszfüzetben megfogalmazott tézisek egyáltalán nem tartalmaznak formulákat, amely egy teljesítményelemzéssel foglalkozó dolgozat esetében szokatlannak mondható), de gondolom a szerzőnek sem ez volt a célja. Az elvi kérdés itt az, hogy vajon lehetett volna-e képletekkel, tételekkel és illusztráló ábrákkal ellátott téziszfüzetet készíteni, amelynek mérete értelmes határokon belül van (pl. nem több mint 20-25 oldal) és egyben megmarad a jelenlegi forma közérthetőségének szintjén. A jelenlegi forma kétségtelen előnye a tömörsége, és az informális megfogalmazásokból adódó, intuícióra alapozó prezentálás (megértetés) lehetősége.

A disszertáció első részében a motivációk ismertetése természetes, a kontribúciók összefoglalása pedig mindenképpen eleget tesz *AZ MTA VI. MŰSZAKI TUDOMÁNYOK OSZTÁLYA Doktori Ügyrendje 1. sz. mellékletének 1. pontja 1. Követelmények a doktori műre és a tézisekre vonatkozóan* azon követelményének, hogy "Az értekezésnek tartalmaznia kell a téziszfüzetben külön összefoglalt téziseket is". (Nyilván az eredmények összefoglalása jobban nézett volna ki a disszertáció végén). Mivel a jelölt munkája a szakirodalomban közismert kvázi születési-halálozási folyamatokhoz, illetve a spektrális felbontáson alapuló numerikus megoldási módszerekhez kötődik, így a bevezető rész rövid második fejezete ezzel a témakörrel foglalkozik. A 2.3. alfejezetben a szerző megemlíti (referenciákkal), hogy számos más módszer van QBD és QBD-M folyamatok megoldására, és ezeket nagyon röviden jellemzi is. Amit hiányoltam itt, hogy a szerző röviden kifejtse a spektrális felbontáson alapuló megoldások várható viszonyát vagy teljesítményét a hivatkozott más megoldásokhoz képest.

A jelölt a disszertáció harmadik részében a többszerveres Markovi sorbanállási rendszerek általánosításaival és ezek alkalmazásaivaé foglalkozik. A jelölt bevezeti az ún. *HetSigma* sorbanállási modellt ($MM\sum_{k=1}^K CPP_k/GE/c/LG$ -queue), amely analízisével kapcsolatos új eredményei a 3.3. alfejezetekben találhatóak. A 3.2. fejezetben a szerző ismerteti a bevezetett modell állandósult állapotaira vonatkozó egyensúlyi egyenletrendszer, amely sajnos végtelen sok egyenletből áll, végtelen sok változót tartalmaz, erre a szakirodalomban sem számítási módszer sem megoldási metodológia eddig nem létezett. A matematikai

kihívás itt az, hogy találjunk hatékony megoldást erre a modellre, amely kezelni tudja a nemszabályos egyensúlyi egyenleteket és az állapotter robbanását. A 3.3. fejezet ismerteti a szerző megoldási módszerét, amely az egyensúlyi egyenletek transzformációin és a transzformált egyenletek tulajdonságainak meghatározásán alapszik. Az egyensúlyi egyenletek transzformációinak tárgyalása meglehetősen tömör és induktív, talán lehetett volna kiegészítésképpen intuitíven is érzékeltetni (vagy néhány sorban informálisan értelmezni) azt, hogy milyen ötletek vezettek az egyenlettranszformációk megalkotására.

Ezzel kapcsolatban azt kérdezem, hogy a HetSigma sorbanállási rendszerrel modellezendő valós fizikai rendszer szempontjából, a megoldásához alkalmazott egyenlettranszformációknak van-e valamilyen konkrét fizikai jelentése?

A 3.3.3 alfejezetben a szerző további "fontos" tulajdonságokat prezentál tételszerűen (természetesen a bizonyításokkal együtt, amelyek véleményem szerint helyesek), főleg a spektrális felbontásban szereplő mátrixokra vonatkozóan.

A fent említett alfejezetekből kissé hiányoltam a megoldási módszer számítási komplexitásának elemzését (a hatékonyságra a szerző utal a 3.4. Összefoglalás alfejezetben). Nevezetesen azt, hogy ha adottak a HetSigma sorbanállási modell paraméterei, akkor hogyan alakul a számítási komplexitása annak az eljárásnak, amely generálja a (3.15) spektrális komponenseket tartalmazó (QBD-M rendszert jellemző) egyenletrendszert.

A HetSigma sorbanállási rendszert több különböző infokommunikációs rendszer hatékony modelljeként ismerhetjük meg a jelölt publikációiból (mint pl. optikai börszt kapcsolat, többprotokollós cimkekapcsolás, UMTS/HSDPA analitikus modelljes). A szerző a disszertációban (a 4. részben) az Apache web kiszolgáló szoftvere modellezését prezentálja. A HTTP kéréseket leíró érkezési folyamatot tekintve a szerző a WAGON sorbanállási modell általánosítását használja (amely [1]-ben kerül ismertetésre), azaz az érkezési folyamat MMCP és az érkezések között eltelt idő pedig MMGE folyamattal van jellemezve. Ez pontosan a 4.3.3.1. alfejezetben kerül ismertetésre, a szerző viszont korábban többször "előre" utal a modellre mint a jól ismert WAGON modell általánosítására. Itt megjegyzem, hogy szerintem jó lett volna röviden leírni, hogy a WAGON modell pontosan milyen folyamatokkal jellemzi az igények érkezését (tehát, hogy mihez képest általánosítás a szerző érkezési modellje).

A web szerver kiszolgáló folyamatainak a száma a működés során változik, függ pl. az aktuálisan kiszolgálás alatt lévő igények és az "üresjáratban" lévő (idle) kiszolgáló folyamatok számától. A szerver az üresjáratú kiszolgáló folyamatok számát egy előre meghatározott tartományban igyekszik tartani (új processzek generálásával illetve meglévő processzek megszüntetésével). A kiszolgáló folyamatok létrehozásai ill megszüntetései között eltelt idők szintén általánosított exponenciális eloszlással vannak modellezve, így kezelhető matematikai modellt biztosítva az elemzésekhez. Ennek a megállapításnak viszont ellentmondani látszik a 49. oldal alján szereplő U_j mennyiség bevezetése, amely azt írja le, hogy a kiszolgáló processzek tranzíciós rátája i_2 és $i_2 + 1$ állapotok között η (amennyiben $i_2 - j < h_{\min}$), és az i_2 és $i_2 - 1$ állapotok között pedig ϵ (amennyiben $i_2 - j > h_{\max}$). Kérem a szerzőt, hogy tisztázza ezt! Nem egyszerűen η és ϵ paraméterű "sima" exponenciális eloszlásokról van szó?

A modell teljesítménykiértékelése két fő részből áll, a validációból és a web szerver által

használt paraméterek (Apache directives) hatásának vizsgálatából. A leírás tömör (nyilván a terjedelmi korlátok miatt is), de lényegretörő, az eredmények imponálóak, még annak fényében is, hogy a szerző egy egyszerűsített érkezési folyamatot (egyállapotú moduláló folyamat) használt. A validációs részben olvashatjuk, hogy a HTTP kiszolgálási folyamat paraméterei mérési statisztikákból lettek származtatva. Korábban (4.3.4. alfejezetben) viszont a fizikai erőforrásokért való versengés kapcsán egy zárt sorbanállási hálózati modellre történik hivatkozás, amely alapján a kiszolgálási ráták meghatározhatók. A disszertációban nem egyértelmű, hogy a felhasznált statisztikák hogyan viszonyulnak a CQN modellből származtatható mennyiségekhez. Kérem a szerzőt, hogy ezt röviden ismertesse! A vonatkozó első tézisben szereplő állításokat új tudományos eredményeknek értékelem.

A disszertáció harmadik része a többszerveres újrapróbálkozási sorbanállási modellekkel és alkalmazásaikkal foglalkozik. Az ötödik fejezet nem kisebb célt, mint a DHCP teljesítőképességének kvantitatív kiértékelését tűzi ki célul.

A DHCP modell elemzéséhez egy olyan QBD modellt alkotott meg a jelölt, amely számítási komplexitása $O(c)$, ellentétben sok más algoritmussal, amelyeknek a futási ideje az IP címek számának köbével ($O(c^3)$ time complexity) arányosan növekszik. Ez véleményem szerint különösen figyelemre méltó eredmény. A DHCP teljesítményének kiértékeléséhez a szerző két mennyiséget vizsgál, a foglalt IP címek átlagos számát és az IP címre várakozó kliensek átlagos számát. A szimulációs vizsgálatokkal kapcsolatban a szerző említi, hogy a szimuláció nem követi az analitikus modell feltételezéseit. A szerző itt arra hivatkozik, hogy a cél ezzel az volt (a 72. oldalon 3 pontban felsorolt különbség), hogy a szimulációs modell minél inkább kövesse a kliensek és a DHCP szerver kölcsönhatását. A kérdéseim erre vonatkozóan a következők:

Az első pont azt mondja, hogy a szimulációs modellben az újrapróbálkozások rátája függ $J(t)$ -től, míg az analitikus modellben fix érték. Mivel a szimulációs és analitikus eredmények jó egyezést mutatnak, következtethetünk-e arra, hogy a vizsgált metrikák nem érzékenyek az újrapróbálkozási ráta állapotfüggésére? Milyen eredményeket kapunk, ha a szimulációban is fix értékre állítjuk az újrapróbálkozási rátát?

Hasonlóan, a harmadik pont szerint egy IP cím bérleti ideje a szimulációban fix érték, az analitikus modellben pedig exponenciális eloszlású. Mi volt az oka, hogy az analitikus modellt "elbonyolítja" a szerző az exponenciális eloszlás alkalmazásával? Miért nem használt fix értéket az analitikus modellben is?

Külön nagyra értékelem a jelölt összehasonlító tesztjeit, melyek segítségével kimutatta, hogy az általa javasolt algoritmus nagyságrendekkel gyorsabban számítja ki a szükséges paramétereket, mint a szakirodalomból ismert eddigi módszerek.

A disszertáció 6. fejezete egy új számítási megközelítést tárgyal vezeték nélküli hálózatban használt védőcsatorna mechanizmus elemzésére. Az ismertett új módszer stabilitás, pontosság és futási idő tekintetében is sokkal jobb az eddig ismert eljárásoknál. A 6.3.3.3. fejezetben a szerző említi, hogy a vonatkozó SE módszer is numerikusan instabil az állapotrobbanás következtében nagy c értékekre, ezért egy új módszert ismertet, amely nem fog ettől a hibától szenvedni. Ezzel kapcsolatban a kérdésem a következő:

A 6.1. algoritmus stabilitása pontosan minek a következménye, illetve lehet-e

általánosítani az algoritmust olyan esetekre, amikor nem csak kettő hanem ennél több nem nulla sajátértéke van a Q mátrixnak ?

Az 5., 6. és 7. fejezetben ismertetett új eredményeket a 2. tézis foglalja össze, amelyeket új tudományos eredményeknek elfogadok.

A 3. téziscsoport virtuális-gép szolgáltatók analízisével foglalkozik. Nagyon innovatívnak találtam a bevezetett új vakációs sorbanállási modellt, mely segítségével elemezhető a karbantartási politika. A szerző a 8.1. tételben megadja a rendszer működésére vonatkozó stabilitási feltételt, amelynek fontos következményeként megállapítható, hogy sem a vakáció periódusa, sem a vakáción lévő szerverek száma nem befolyásolja a rendszer stabilitását. Ezt nagyon fontos elméleti eredménynek tartom, amelynek közvetlen gyakorlati hatása van a hálózatüzemeltetők karbantartási tevékenységére. A vonatkozó 3. tézist új tudományos eredménynek értékelem.

Összefoglaló értékelés

Do Van Tien véleményem szerint teljesítette az MTA doktora címmel szemben támasztott követelményeket. A jelölt habitusa, folyamatos magas színvonalú publikációs tevékenysége és a disszertációban ismertetett új tudományos eredményei alapján javaslom részére az MTA doktora cím odaítélését.

Budapest, 2011. 04. 13.

Bíró József
egyetemi tanár, az MTA doktora
Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék

Hivatkozások

- [1] Xue Liu, Lui Sha, Yixin Diao, Steven Froehlich, Joseph L. Hellerstein, and Sujay Parekh. Online Response Time Optimization of Apache Web Server. In Eleventh International Workshop on Quality of Service (IWQoS 2003), pages 46–478, 2003.